UCLA
UCLA Previously Published Works

Title

D-optimal designs formulti-response linear mixed models

Permalink
https://escholarship.org/uc/item/74k7w4tz

Authors
Liu, Xin
Yue, Rong-Xian
Wong, Weng Kee

Publication Date
2018-09-08

DOI
10.1007/s00184-018-0679-7

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Metrika (2019) 82:87–98
https://doi.org/10.1007/s00184-018-0679-7

D-optimal designs for multi-response linear mixed models

Xin Liu[1] · Rong-Xian Yue[2] · Weng Kee Wong[3]

Received: 26 February 2018 / Published online: 8 September 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Linear mixed models have become popular in many statistical applications during recent years. However design issues for multi-response linear mixed models are rarely discussed. The main purpose of this paper is to investigate D-optimal designs for multi-response linear mixed models. We provide two equivalence theorems to characterize the optimal designs for the estimation of the fixed effects and the prediction of random effects, respectively. Two examples of the D-optimal designs for multi-response linear mixed models are given for illustration.

Keywords D-optimal designs · Multi-response · Linear mixed model · Equivalence theorem

1 Introduction

The mixed effects model (or, more simply, mixed model) is a popular choice to analyze correlated data such as longitudinal or repeated measurement and panel data, see Laird and Ware (1982), Baltagi (1995) and Diggle et al. (2002). The properties of mixed models have been well studied in the literature and detailed descriptions of the analysis of these models can be found in the books of Davidian and Giltinan (1995), Vonesh and Chinchilli (1997) and Verbeke and Molenberghs (2000), among others.

In many applications, the response variable has multiple random components that may interact with each other. To make inference, it is necessary to jointly model these multiple components. Such models are called multi-response mixed effects models and they are increasingly used to analyze different types of data. For example, Dahm et al. (1983) used multi-response mixed models to analyze animal breeding experiments to draw inference for an underlying genotype covariance matrix. Sun et al. (2003)

✉ Rong-Xian Yue
yue2@shnu.edu.cn

1 College of Science, Donghua University, Shanghai 201600, China

2 Department of Mathematics, Shanghai Normal University, Shanghai 200234, China

3 Department of Biostatistics, University of California, Los Angeles, CA 90095-1772, USA

proposed bivariate mixed effects models to study growth curves for children born as singletons in the city of Uppsala during 1973–1977, Verbeke and Molenberghs (2000, Section 24.1) employed multi-response linear mixed effects models to analyze the experimental data of systolic and diastolic blood pressure and Jensen et al. (2012) applied multi-response linear mixed effects modeling to analyze experimental data on the relation between air quality and the performance of office work.

The analysis of these models depends on the experimental design. When the experimental settings are under the control of the investigator, design issues must be carefully addressed to attain maximal accuracy of the statistical inference at minimal cost. Design issues in mixed effects models with univariate response have been considered in the literature. Fedorov and Hackl (1997) derived an equivalence theorem to confirm D-optimality of an approximate design for a random coefficient regression model. Entholzner et al. (2005) obtained optimal and efficient designs under mixed models. Schmelter (2007a) showed that optimal designs for linear mixed models could be restricted to the class of group-wise identical individual designs. Schmelter (2007b) concluded that optimal designs in the class of single-group designs remain optimal in the larger class having more group designs. Schwabe and Schmelter (2008), Schmelter et al. (2007) and Luoma et al. (2007) investigated optimal designs under random intercept models, random slope models and random coefficient cubic regression models, respectively. Debusho and Haines (2008, 2011) discussed the construction of V- and D-optimal population designs for linear and quadratic regression models with a random intercept term. Recently, Prus and Schwabe (2016) studied optimal designs for the prediction of individual parameters in hierarchical models.

The design problem for multi-response mixed effects models has received little attention to date. In this paper we are interested to find optimal experimental designs for estimation of the fixed effects (population parameters) and prediction of the random effects (individual parameters) in a multi-response linear mixed model such as a multi-response random coefficient regression model.

Our aim is to provide a multi-response version of the equivalence theorem for D-optimality given in Prus and Schwabe (2016). Section 2 introduces the multi-response linear mixed model and provides preliminaries. Section 3 considers multi-response linear mixed models and provides equivalence theorems for the D-optimality with respect to estimation of the fixed effects and the prediction of random effects, respectively. We also present two examples of the D-optimal designs for multi-response mixed models in Sect. 3. Proofs are deferred to Appendix.

2 Model specification and preliminaries

Throughout, we suppose that each observation is a $r \times 1$ vector of responses and we denote the jth multi-response from the ith subject by the vector y_{ij}. Our multi-response random coefficient regression model of interest is defined on a user-defined compact design space $\mathcal{X}$ and is given by

$$\boldsymbol{y}_{ij} = \boldsymbol{F}(x_{ij})\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \tag{1}$$

where the jth observation from individual i is taken at the experimental setting $x_{ij} \in \mathcal{X}$, n is the number of individuals, m_i is the number of observations from individual i, $\boldsymbol{F}$ is the $r \times p$ matrix of known regression functions, $\boldsymbol{\varepsilon}_{ij}$ is a vector of random errors having mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The individual parameters $\boldsymbol{\beta}_i$ are assumed to be random vectors with mean $\mathrm{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ and covariance matrix $\mathrm{Cov}(\boldsymbol{\beta}_i) = \boldsymbol{D}$. It is assumed that $\boldsymbol{\Sigma}$ is positive-definite, and $\boldsymbol{D}$ is positive semidefinite with rank $q \leq p$. Further we suppose that all individual parameters $\boldsymbol{\beta}_i$'s and all observational errors $\boldsymbol{\varepsilon}_{ij}$'s are uncorrelated. Note that $\boldsymbol{D}$ can be singular which allows for some individual parameters to be non-random.

Alternatively, by separating the random effects from the population mean, the model (1) can be rewritten as

$$\boldsymbol{y}_{ij} = \boldsymbol{F}(x_{ij})\boldsymbol{\beta} + \boldsymbol{F}(x_{ij})\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \tag{2}$$

where $\boldsymbol{b}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}$ is the individual effect compared with the population mean. In both models (1) and (2), $\boldsymbol{y}_{ij}$'s are random vectors with covariance matrix $\mathrm{Cov}(\boldsymbol{y}_{ij}) = \boldsymbol{F}(x_{ij})\boldsymbol{D}\boldsymbol{F}^T(x_{ij}) + \boldsymbol{\Sigma}$. The observations from the same individual are correlated with covariance structure $\mathrm{Cov}(\boldsymbol{y}_{ij}, \boldsymbol{y}_{ik}) = \boldsymbol{F}(x_{ij})\boldsymbol{D}\boldsymbol{F}^T(x_{ik})$, $j \neq k$, and observation from different individuals are uncorrelated. If we let $\boldsymbol{Y}_i = (\boldsymbol{y}_{i1}^T, \ldots, \boldsymbol{y}_{im_i}^T)^T$, it is instructive to write the model in matrix form as

$$\boldsymbol{Y}_i = \boldsymbol{F}_i\boldsymbol{\beta} + \boldsymbol{F}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{3}$$

where we assume the design matrix $\boldsymbol{F}_i = (\boldsymbol{F}^T(x_{i1}), \ldots, \boldsymbol{F}^T(x_{im_i}))^T$ for individual i has full column rank and $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}^T, \ldots, \boldsymbol{\varepsilon}_{im_i}^T)^T$ is the corresponding vector of observational errors with covariance matrix $\boldsymbol{I}_{m_i} \otimes \boldsymbol{\Sigma}$. Then $\mathrm{E}(\boldsymbol{Y}_i) = \boldsymbol{F}_i\boldsymbol{\beta}$ and $\mathrm{Cov}(\boldsymbol{Y}_i) = \boldsymbol{F}_i\boldsymbol{D}\boldsymbol{F}_i^T + \boldsymbol{I}_{m_i} \otimes \boldsymbol{\Sigma}$, where $\boldsymbol{I}_k$ is the $k \times k$ identity matrix.

The full vector $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T)^T$ of the observations of all individuals can be expressed in matrix form as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{X} = (\boldsymbol{F}_1^T, \ldots, \boldsymbol{F}_n^T)^T$, $\boldsymbol{Z} = \mathrm{diag}(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n)$, $\boldsymbol{b} = (\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_n^T)^T$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^T, \ldots, \boldsymbol{\varepsilon}_n^T)^T$. It follows that the covariance matrix of the observational vector $\boldsymbol{Y}$ is block diagonal with

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{Y}) &= \mathrm{diag}(\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{F}_1^T + \boldsymbol{I}_{m_1} \otimes \boldsymbol{\Sigma}, \ldots, \boldsymbol{F}_n\boldsymbol{D}\boldsymbol{F}_n^T + \boldsymbol{I}_{m_n} \otimes \boldsymbol{\Sigma}) \\
&= \boldsymbol{Z}(\boldsymbol{I}_n \otimes \boldsymbol{D})\boldsymbol{Z}^T + \boldsymbol{I}_N \otimes \boldsymbol{\Sigma},
\end{aligned}$$

where $N = m_1 + \cdots + m_n$ is the total number of observations. In what is to follow, we report the estimators when the matrix $\boldsymbol{D}$ is non-singular and provide some details for the derivation when $\boldsymbol{D}$ is singular.

First, we consider the case that the covariance matrix $\boldsymbol{D}$ is non-singular. According to Henderson et al. (1959) and Christensen (2002), the best linear unbiased estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and the best linear unbiased prediction $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ are given by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Y} \end{pmatrix}, \tag{5}$$

where $\boldsymbol{R} = \mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}$ and $\boldsymbol{G} = \mathrm{Cov}(\boldsymbol{b}) = \boldsymbol{I}_n \otimes \boldsymbol{D}$. Moreover, Henderson (1975) showed that the joint covariance matrix of both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}} - \boldsymbol{b}$ is

$$\mathrm{Cov}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{b}} - \boldsymbol{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix}^{-1}. \tag{6}$$

It follows from (5) and (6) that

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R})^{-1}\boldsymbol{Y}$$

and

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R})^{-1}\boldsymbol{X})^{-1} = \left(\sum_{i=1}^{n} \boldsymbol{F}_i^T(\boldsymbol{F}_i\boldsymbol{D}\boldsymbol{F}_i^T + \boldsymbol{I}_{m_i} \otimes \boldsymbol{\Sigma})^{-1}\boldsymbol{F}_i\right)^{-1}.$$

In the case when $\mathrm{rank}(\boldsymbol{D}) = q < p$, there is a $p \times q$ matrix $\boldsymbol{K}$ with $\boldsymbol{D} = \boldsymbol{K}\boldsymbol{K}^T$ and $\mathrm{rank}(\boldsymbol{K}) = q$ such that $\boldsymbol{K}^T\boldsymbol{K}$ is non-singular. As it has done in Prus and Schwabe (2016), the model (2) can be written as

$$\boldsymbol{y}_{ij} = \boldsymbol{F}(x_{ij})\boldsymbol{\beta} + \boldsymbol{F}(x_{ij})\boldsymbol{K}\boldsymbol{c}_i + \boldsymbol{\varepsilon}_{ij},$$

where $\boldsymbol{c}_i = (\boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{K}^T(\boldsymbol{\beta}_i - \boldsymbol{\beta})$ are random effects. Then the complete observation vector can be expressed as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}\boldsymbol{c} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{X} = (\boldsymbol{F}_1^T, \ldots, \boldsymbol{F}_n^T)^T$, $\tilde{\boldsymbol{Z}} = \mathrm{diag}(\boldsymbol{F}_1\boldsymbol{K}, \ldots, \boldsymbol{F}_n\boldsymbol{K})$ and $\boldsymbol{c} = (\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_n^T)^T$, which results in a model equation with non-singular covariance matrices $\boldsymbol{R} = \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{G}} = \mathrm{Cov}(\boldsymbol{c}) = \boldsymbol{I}_n \otimes \boldsymbol{I}_q$ respectively. With this notation the joint covariance matrix of both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{c}} - \boldsymbol{c}$ is

$$\mathrm{Cov}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{c}} - \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\tilde{\boldsymbol{Z}} \\ \tilde{\boldsymbol{Z}}^T\boldsymbol{R}^{-1}\tilde{\boldsymbol{Z}} & \tilde{\boldsymbol{Z}}^T\boldsymbol{R}^{-1}\tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{G}}^{-1} \end{pmatrix}^{-1}, \tag{7}$$

which is similar to that in (6).

There are practical situations where physical constraints may force the experimenter to observe all individuals under the same regime, i.e., all individuals i have the same number $m_i = m$ of observations at the same values $x_{ij} = x_j$ of the experimental settings, which is called a balanced design [or an identical individual design in Prus and Schwabe (2016)]. For balanced designs, the covariance matrix of $\hat{\boldsymbol{\beta}}$ simplifies to

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\left(\left(\sum_{i=1}^{m} \boldsymbol{F}^T(x_i)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x_i)\right)^{-1} + \boldsymbol{D}\right). \tag{8}$$

Let $\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{b}}_i$ be the predictor of the individual parameters $\boldsymbol{\beta}_i$ and let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1^T, \ldots, \hat{\boldsymbol{\beta}}_n^T)^T$ be the predictor for $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_n^T)^T$ of all individual coefficients. Then for balanced designs using similar arguments in Prus and Schwabe (2016), it can be shown that the mean-squared error matrix (MSE-matrix) of $\hat{\boldsymbol{\theta}}$ is

$$\begin{aligned}\mathrm{MSE}(\hat{\boldsymbol{\theta}}) = & -\frac{1}{n}\boldsymbol{J}_n \otimes \left(\sum_{i=1}^{m} \boldsymbol{F}^T(x_i)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x_i)\right)^{-1} \\ & + \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right) \otimes \left(\boldsymbol{D} - \boldsymbol{D}\left(\left(\sum_{i=1}^{m} \boldsymbol{F}^T(x_i)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x_i)\right)^{-1} + \boldsymbol{D}\right)^{-1}\boldsymbol{D}\right),\end{aligned} \tag{9}$$

where $\boldsymbol{J}_n$ is the $n \times n$ matrix with all entries equal to 1. In the case that $\boldsymbol{D}$ is non-singular, the MSE-matrix of $\hat{\boldsymbol{\theta}}$ simplifies to

$$\begin{aligned}\mathrm{MSE}(\hat{\boldsymbol{\theta}}) = & -\frac{1}{n}\boldsymbol{J}_n \otimes \left(\sum_{i=1}^{m} \boldsymbol{F}^T(x_i)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x_i)\right)^{-1} \\ & + \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right) \otimes \left(\sum_{i=1}^{m} \boldsymbol{F}^T(x_i)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x_i) + \boldsymbol{D}^{-1}\right)^{-1}.\end{aligned}$$

3 Optimal designs

Our objective is to efficiently estimate the fixed effects $\boldsymbol{\beta}$ based on the covariance matrix $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$, or to accurately predict the random effects $\boldsymbol{\beta}_i$'s based on the MSE-matrix, $\mathrm{MSE}(\hat{\boldsymbol{\theta}})$. Clearly, both these matrices depend on the design of the experiment matrix, i.e., on the choice of the experimental settings x_{ij}'s. For this purpose, we wish to chose experimental settings such that the covariance matrix or the MSE-matrix is smallest in some way. One common way to achieve this goal is to construct *D*-optimal designs.

We focus on designs and suppose $x_1, \ldots, x_k$ are the support points and they are replicated $n_1, \ldots, n_k$ times, respectively. We represented such a design by

$$\xi = \begin{Bmatrix} x_1 & \ldots & x_k \\ n_1 & \ldots & n_k \end{Bmatrix}.$$

To further simplify our presentation, we consider approximate designs in the sense of Kiefer (1959) and drop the requirement that the replication numbers n_l have to be positive integers; only the conditions $n_l \geq 0$ and $\sum_{l=1}^{k} n_l = m$ must be satisfied for each individual. In particular, for approximate designs, only $k, x_1, \ldots, x_k$ and the proportion of observations to be taken at each x_i have to be determined.

The normalized Fisher information matrix of an approximate design ξ under multiresponse linear models is defined by

$$M(\xi) = \int_{\mathcal{X}} F^{\top}(x)\Sigma^{-1}F(x)d\xi(x),$$

see Section 1.7 in Fedorov (1972). Accordingly, for any approximate design ξ the standardized individual information matrix is defined by

$$\boldsymbol{M}(\xi) = \frac{1}{m}\sum_{l=1}^{k} n_l \boldsymbol{F}^T(x_l)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x_l),$$

see a similar consideration in Prus and Schwabe (2016). The matrix $\boldsymbol{M}(\xi)$ stands for the information obtained per observation and $m\boldsymbol{M}(\xi)$ corresponds to the information contributed by the observations at the experimental settings of per individual. Denote by Ξ the set of all approximate designs with non-singular information matrix on $\mathcal{X}$.

With this notation we can express the covariance matrix of $\hat{\boldsymbol{\beta}}$ given in (8) and the MSE-matrix of $\hat{\boldsymbol{\theta}}$ given in (9) corresponding to an approximate design ξ by

$$\mathrm{Cov}_{\xi}(\hat{\boldsymbol{\beta}}) = \frac{1}{nm}(\boldsymbol{M}^{-1}(\xi) + \boldsymbol{\Delta}) \tag{10}$$

and

$$\mathrm{MSE}_{\xi}(\hat{\boldsymbol{\theta}}) = \frac{1}{m}\left\{-\frac{1}{n}\boldsymbol{J}_n \otimes \boldsymbol{M}^{-1}(\xi) + \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right) \otimes (\boldsymbol{\Delta} - \boldsymbol{\Delta}(\boldsymbol{M}^{-1}(\xi) + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\right\}, \tag{11}$$

where $\boldsymbol{\Delta} = m\boldsymbol{D}$. When $\boldsymbol{D}$ is non-singular, this expression simplifies to

$$\mathrm{MSE}_{\xi}(\hat{\boldsymbol{\theta}}) = \frac{1}{m}\left\{-\frac{1}{n}\boldsymbol{J}_n \otimes \boldsymbol{M}^{-1}(\xi) + \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right) \otimes (\boldsymbol{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}\right\}.$$

3.1 *D*-optimal design for the estimation of fixed effects

First, we consider D-optimal designs for the estimation of fixed effects $\boldsymbol{\beta}$. A design is called D-optimal for the estimation of fixed effects $\boldsymbol{\beta}$ if it minimizes

$$\psi_D(\xi) = \ln|\mathrm{Cov}_{\xi}(\hat{\boldsymbol{\beta}})|, \tag{12}$$

where the matrix $\mathrm{Cov}_\xi(\hat{\boldsymbol{\beta}})$ is given in (10). The *D*-optimal designs can be characterized by the following equivalence theorem.

Theorem 1 *Let*

$$\phi(x,\xi) = tr\Big\{\boldsymbol{F}^T(x)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x)\boldsymbol{M}^{-1}(\xi)(\boldsymbol{M}^{-1}(\xi)+\boldsymbol{\Delta})^{-1}\boldsymbol{M}^{-1}(\xi)\Big\}.$$

Then a design $\xi^* \in \Xi$ is D-optimal for the estimation of the fixed effects $\boldsymbol{\beta}$ if and only if

$$\sup_{x\in\mathcal{X}} \phi(x,\xi^*) = tr\left\{(\boldsymbol{M}^{-1}(\xi^*)+\boldsymbol{\Delta})^{-1}\boldsymbol{M}^{-1}(\xi^*)\right\}.$$

Moreover, the supremum over $\mathcal{X}$ is achieved at the support points of ξ^*.

Example 1 Design of circular measurements.

The circular feature in a mechanical object is one of the most basic geometric primitives. Its specification can be described easily by a center and a radius. Due to imperfections introduced at the manufacturing stage, the desired feature may not be truly circular. In order to control the production, we need to estimate the geometric parameters (center and radius), which requires data on machined parts along their circumferences, and a corresponding statistical model. In practice the data can be obtained using a Coordinate Measuring Machine, while one of the models adopted for such data is the mixed effects model provided by Wang and Lam (1997), which also considers the variability in the center locations of different machined parts.

Let (y_{1ij}, y_{2ij}) be the jth measurement of the coordinates on the circumference of the ith machined part, $(a_1+u_{1i}, a_2+u_{2i})^T$ be the location of the center of part i, $i=1,\ldots,n$, $j=1,\ldots,m$, where n is the number of machined parts, m is the number of measurements taken from the circumference, $(a_1,a_2)^T$ is the designed location of center of the part and $(u_{1i},u_{2i})^T$ is the random drift in the location of the center of the ith machined part. Moreover, the radius of part i, ρ_i is fixed but unknown. Let $\tau_{i(j)}$ be the measured angle of the jth measurement point on the ith part. Because measurements are all taken with respect to some fixed but usually unknown direction, the angular difference between measurements, $\tau_{i(j+1)}-\tau_{i(j)}$, are controlled by the investigator and assumed to be identical for all machined parts. Write

$$\tau_{i(j)} = \theta_{i0}+\theta_j, \quad j=1,2,\ldots,m,$$

$\alpha_i=\rho_i\cos\theta_{i0}$ and $\beta_i=\rho_i\sin\theta_{i0}$, where θ_j is known and θ_{i0} is fixed but unknown. The design space is $\mathcal{X}=[-\alpha/2,\alpha/2]$ and consists of all possible values of θ_j. In practice, the value of α is user-selected and $\alpha\in(0,2\pi]$ is the length of the arc. Then for the data points (y_{1ij},y_{2ij}) are used to fit the following model (see Wang and Lam 1997)

$$\begin{aligned} y_{1ij} &= a_1+\alpha_i\cos\theta_j-\beta_i\sin\theta_j+u_{1i}+\varepsilon_{1ij},\\ y_{2ij} &= a_2+\alpha_i\sin\theta_j+\beta_i\cos\theta_j+u_{2i}+\varepsilon_{2ij}, \end{aligned} \quad i=1,\ldots,n,\quad j=1,\ldots,m, \tag{13}$$

where $u_{ki} \sim N(0, \sigma_0^2)$, $\varepsilon_{kij} \sim N(0, \sigma^2)$, $k = 1, 2$, and all u_{ki}'s and ε_{kij}'s are independent.

For this model, the matrix of regression functions is

$$\boldsymbol{F}(\theta) = (\boldsymbol{I}_2, \boldsymbol{e}_i^T \otimes \boldsymbol{A}(\theta)), \quad \text{where } \boldsymbol{A}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

and $\boldsymbol{e}_i$ is the ith unit vector in $\mathbb{R}^n$. The population parameter vector that we wish to estimate is

$$\boldsymbol{\beta} = (a_1, a_2, \alpha_1, \beta_1, \ldots, \alpha_n, \beta_n)^T.$$

Let $c(\xi) = \frac{1}{m}\sum_{l=1}^{k} n_l \cos\theta_l$ and $s(\xi) = \frac{1}{m}\sum_{l=1}^{k} n_l \sin\theta_l$. By Theorem 1, we can verify that the symmetric design

$$\xi_D^* = \begin{Bmatrix} -\theta_* & \theta_* \\ m/2 & m/2 \end{Bmatrix}$$

is D-optimal, where $\theta_* = \min(\alpha/2, \pi/2)$. Further, we have $c(\xi_D^*) = \cos\theta_*$, $s(\xi_D^*) = 0$ and

$$\mathrm{Cov}_{\xi_D^*}(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\begin{pmatrix} (s_1 + \sigma_0^2)\boldsymbol{I}_2 & -s_1\cos\theta_* \mathbf{1}_n^T \otimes \boldsymbol{I}_2 \\ -s_1\cos\theta_* \mathbf{1}_n \otimes \boldsymbol{I}_2 & ns_2\boldsymbol{I}_{2n} + (s_1 - s_2)\mathbf{1}_n\mathbf{1}_n^T \otimes \boldsymbol{I}_2 \end{pmatrix},$$

where

$$s_1 = \frac{\sigma^2}{1 - \cos^2\theta_*}, \quad s_2 = \frac{(m\sigma_0^2 + \sigma^2)\sigma^2}{m\sigma_0^2(1 - \cos^2\theta_*) + \sigma^2}.$$

Straightforward calculation shows that

$$\begin{aligned}\phi(\theta, \xi_D^*) = &-\frac{4\cos\theta_*}{1 - \cos^2\theta_*}\cos(\theta) + \frac{4\cos^2\theta_*}{1 - \cos^2\theta_*} \\ &+ \frac{2(n-1)m\sigma_0^2\cos^2\theta_*}{m\sigma_0^2(1 - \cos^2\theta_*) + \sigma^2} + \frac{2\sigma^2}{m\sigma_0^2 + \sigma^2},\end{aligned}$$

for all $\theta \in \mathcal{X} = [-\alpha/2, \alpha/2]$ and

$$\mathrm{tr}\left\{(\boldsymbol{M}^{-1}(\xi_D^*) + \boldsymbol{\Delta})^{-1}\boldsymbol{M}^{-1}(\xi_D^*)\right\} = 2n + \frac{2\sigma^2}{m\sigma_0^2 + \sigma^2}.$$

Since $\phi(\theta, \xi_D^*)$ decreases in $\cos(\theta)$ when $\alpha < \pi$, it is easy to see that

$$\sup_{\theta \in \mathcal{X}} \phi(\theta, \xi_D^*) = 2n + \frac{2\sigma^2}{m\sigma_0^2 + \sigma^2}.$$

It follows from Theorem 1 that ξ_D^* is a *D*-optimal design.

3.2 *D*-optimal design for the prediction of random effects

We now consider optimal designs for the prediction of the random effects $\boldsymbol{\beta}_i$'s. Since the accuracy of prediction is measured by the MSE-matrix, one may adopt the *D*-criterion by minimizing the determinant of the MSE-matrix:

$$\psi_D^{pred}(\xi) = \ln |\mathrm{MSE}_\xi(\hat{\boldsymbol{\theta}})|, \tag{14}$$

where the matrix $\mathrm{MSE}_\xi(\hat{\boldsymbol{\theta}})$ is given in (11). However, as pointed out by Prus and Schwabe (2016), this makes sense when the matrix $\boldsymbol{D}$ is non-singular, otherwise the determinant of the MSE-matrix is zero. Consequently, Prus and Schwabe (2016) proposed a modified *D*-criterion for prediction in terms of the logarithm of the product of the $(n-1)q + p$ largest eigenvalues of the MSE-matrix as follows:

$$\psi_D^{pred}(\xi) = \ln |\boldsymbol{M}^{-1}(\xi)| + (n-1)\ln \sum_{l=1}^{q} \lambda_l(\xi, \boldsymbol{\Delta}), \tag{15}$$

where $\lambda_l(\xi, \boldsymbol{\Delta}), l = 1, \ldots, q$, are the q largest eigenvalues of $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\boldsymbol{M}^{-1}(\xi) + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$. In the regular case ($q = p$), the definition in (15) becomes (14).

With this definition a *D*-optimal design can be characterized by the following equivalence theorem.

Theorem 2 *Let*

$$\begin{aligned}\phi(x, \xi) = tr\left\{\boldsymbol{F}^T(x)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x)\boldsymbol{M}^{-1}(\xi)\right\} \\ +(n-1)tr\left\{\boldsymbol{F}^T(x)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x)(\boldsymbol{\Delta} - \boldsymbol{\Delta}(\boldsymbol{M}^{-1}(\xi) + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\right\}.\end{aligned}$$

Then a design $\xi^* \in \Xi$ is D-optimal for the prediction of the random effects $\boldsymbol{\beta}_i$'s if and only if

$$\sup_{x \in \mathcal{X}} \phi(x, \xi^*) = p + (n-1)tr\left\{(\boldsymbol{M}^{-1}(\xi^*) + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\right\}.$$

Moreover, the supremum over $\mathcal{X}$ is achieved at the support points of ξ^*.

Corollary 3 *If $\boldsymbol{D}$ is non-singular, let*

$$\begin{aligned}\phi(x, \xi) = tr\left\{\boldsymbol{F}^T(x)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x)\boldsymbol{M}^{-1}(\xi)\right\} \\ +(n-1)tr\left\{\boldsymbol{F}^T(x)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x)(\boldsymbol{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}\right\}.\end{aligned}$$

Then a design $\xi^* \in \Xi$ is D-optimal for the prediction of the random effects $\boldsymbol{\beta}_i$'s if and only if

$$\sup_{x\in\mathcal{X}} \phi(x,\xi^*) = p + (n-1)tr\left\{(\boldsymbol{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1}\boldsymbol{M}(\xi^*)\right\}.$$

Moreover, the supremum over $\mathcal{X}$ is achieved at the support points of ξ^*.

Example 2 Linear regression with random intercept terms.

In recent years, indoor environment effects on the performance of office work have been studied extensively. Mixed-effects modelling has been effectively employed in the analysis of the experimental results. If the different components of the outcome variable are not independent, a multi-response linear mixed model can be used to explore the correlations among the response variables. For example, the following model was employed by Jensen et al. (2012) to analyze experimental data on the relation between air quality and the performance of office work.

$$\begin{aligned} y_{1ij} &= \mu_1 + \theta_1 x_{ij} + P_{1i} + \varepsilon_{1ij}, \\ y_{2ij} &= \mu_2 + \theta_2 x_{ij} + P_{2i} + \varepsilon_{2ij}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,m, \\ y_{3ij} &= \mu_3 + \theta_2 x_{ij} + P_{3i} + \varepsilon_{3ij}, \end{aligned} \tag{16}$$

where $P_i = (P_{1i}, P_{2i}, P_{3i})^T$ are the individual random effects having a normal distribution $N_3(\mathbf{0}, \boldsymbol{\Sigma}_P)$, $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{1ij}, \varepsilon_{2ij}, \varepsilon_{3ij})^T \sim N_3(\mathbf{0}, \boldsymbol{\Sigma})$, and all P_i's and ε_{ij}'s are independent.

We consider the optimal design problem for model (16). The experimental region is taken as $\mathcal{X} = [0, 1]$. We show using Theorem 2 that the D-optimal design for the prediction of the individual random effects is given by

$$\xi_D^* = \begin{Bmatrix} 0 & 1 \\ m/2 & m/2 \end{Bmatrix}.$$

The matrix of regression functions in model (16) is $\boldsymbol{F}(x) = (1, x) \otimes \boldsymbol{I}_3$, the population parameter is $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3, \theta_1, \theta_2, \theta_3)^T$, and the dispersion matrix is $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{\Sigma}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. The information matrix of ξ_D^* is given by $\boldsymbol{M}(\xi_D^*) = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \otimes \boldsymbol{\Sigma}^{-1}$ and it is easy to obtain that

$$\boldsymbol{F}^T(x)\boldsymbol{\Sigma}^{-1}\boldsymbol{F}(x) = \left((1,x)(1,x)^T\right) \otimes \boldsymbol{\Sigma}^{-1},$$

$$(\boldsymbol{M}^{-1}(\xi_D^*) + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta} = \begin{pmatrix} m(\boldsymbol{\Sigma} + m\boldsymbol{\Sigma}_P)^{-1}\boldsymbol{\Sigma}_P & \mathbf{0} \\ \frac{m}{2}(\boldsymbol{\Sigma} + m\boldsymbol{\Sigma}_P)^{-1}\boldsymbol{\Sigma}_P & \mathbf{0} \end{pmatrix},$$

and

$$\boldsymbol{\Delta} - \boldsymbol{\Delta}(\boldsymbol{M}^{-1}(\xi_D^*) + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta} = \begin{pmatrix} m\boldsymbol{\Sigma}_P - m^2\boldsymbol{\Sigma}_P(\boldsymbol{\Sigma} + m\boldsymbol{\Sigma}_P)^{-1}\boldsymbol{\Sigma}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

It follows that

$$\begin{aligned}\phi(x,\xi_D^*) &= 6(1-2x+2x^2)+(n-1)\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}(m\boldsymbol{\Sigma}_P - m^2\boldsymbol{\Sigma}_P(\boldsymbol{\Sigma}+m\boldsymbol{\Sigma}_P)^{-1}\boldsymbol{\Sigma}_P)\right\}\\ &= 6(1-2x+2x^2)+(n-1)\mathrm{tr}\left\{(\boldsymbol{\Sigma}+m\boldsymbol{\Sigma}_P)^{-1}m\boldsymbol{\Sigma}_P\right\}\\ &= 6(1-2x+2x^2)+(n-1)\mathrm{tr}\left\{(\boldsymbol{M}^{-1}(\xi_D^*)+\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\right\}.\end{aligned}$$

It is clear that $\phi(x,\xi_D^*)$ is nonnegative for any $x\in[0,1]$, and attains its maximum $6+(n-1)\mathrm{tr}\left\{(\boldsymbol{M}^{-1}(\xi_D^*)+\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\right\} = p+(n-1)\mathrm{tr}\left\{(\boldsymbol{M}^{-1}(\xi_D^*)+\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\right\}$ at $x\in\{0,1\}$, the support points of ξ_D^*. It follows from Theorem 2 that the design ξ_D^* is D-optimal over Ξ.

Acknowledgements Dr. Yue and Dr. Liu were partially supported by the National Natural Science Foundation of China under Grants 11471216 and 11871143. Dr. Wong was partially supported by a Grant from the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix

We provide justifications for Theorems 1 and 2 using the same notation from the main text.

A.1. Proof of Theorem 1 and 2

Here we establish the proof using similar arguments in Prus and Schwabe (2016) and invoking the general equivalence theorem (Kiefer 1959).

For the D-optimal criterion $\psi_D(\xi)$ defined in (12), a direct calculation shows that the directional derivative at $\boldsymbol{M}_1$ in the direction of $\boldsymbol{M}_2$ is

$$-\mathrm{tr}\left\{(\boldsymbol{M}_1^{-1}+\boldsymbol{\Delta})^{-1}\boldsymbol{M}_1^{-1}(\boldsymbol{M}_2-\boldsymbol{M}_1)\boldsymbol{M}_1^{-1}\right\}.$$

As in Silvey (1980), a standard argument based on convex analysis then establishes Theorem 1.

To determine the directional derivative for the modified D-criterion $\psi_D^{pred}(\xi)$ we write the criterion in the form

$$\psi_D^{pred}(\xi)=\ln|\boldsymbol{M}^{-1}(\xi)|+(n-1)\left(\ln|m\boldsymbol{K}^T\boldsymbol{K}|+\ln|(m\boldsymbol{K}^T\boldsymbol{M}(\xi)\boldsymbol{K}+\boldsymbol{I}_q)^{-1}|\right),$$

where $m\boldsymbol{K}\boldsymbol{K}^T=\boldsymbol{\Delta}$. As a consequence, this criterion can be identified as a compound criterion with the directional derivative:

$$-\mathrm{tr}\left\{(n-1)(\boldsymbol{\Delta}-\boldsymbol{\Delta}(\boldsymbol{M}_1^{-1}+\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})(\boldsymbol{M}_2-\boldsymbol{M}_1)+\boldsymbol{M}_1^{-1}(\boldsymbol{M}_2-\boldsymbol{M}_1)\right\},$$

which proves Theorem 2. □

References

Baltagi BH (1995) Econometric analysis of panel data. Wiley, New York

Christensen R (2002) Plane answers to complex questions: the theory of linear models. Springer, New York

Dahm PF, Melton BE, Fuller WA (1983) Generalized least squares estimation of a genotypic covariance matrix. Biometrics 39:587–597

Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman and Hall, London

Debusho LK, Haines LM (2008) V-optimal and D-optimal population designs for the simple linear regression model with a random intercept term. J Stat Plan Inference 138:1116–1130

Debusho LK, Haines LM (2011) D- and V-optimal population designs for the quadratic regression model with a random intercept term. J Stat Plan Inference 141:889–898

Diggle PJ, Heagerty P, Liang KE, Zeger SL (2002) Analysis of longitudinal data. Oxford Science, New York

Entholzner M, Benda N, Schmelter T, Schwabe R (2005) A note on designs for estimating population parameter. Listy Biom Biom Lett 42:25–41

Fedorov VV (1972) Theory of optimal experiments. Academic Press, New York

Fedorov VV, Hackl P (1997) Model-oriented design of experiments. Springer, New York

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–477

Henderson CR, Kempthorne O, Searle SR, von Krosigk CM (1959) The estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192–218

Kiefer J (1959) Optimum experimental designs. J R Stat Soc B 21:272–304

Jensen KL, Spiild D, Toftum J (2012) Implementation of multivariate linear mixed-effects models in the analysis of indoor climate performance experiments. Int J Biometeorol 56:129–136

Laird NM, Ware JH (1982) Random effects models for longitudinal data. Biometrics 38:963–974

Luoma A, Nummi T, Sinha BK (2007) Optimal designs in random coefficient cubic regression models. J Stat Plan Inference 137:3611–3617

Prus M, Schwabe R (2016) Optimal designs for the prediction of individual parameters in hierarchical models. J R Stat Soc B 78:175–191

Schmelter T (2007a) The optimality of single-group designs for certain mixed models. Metrika 65:183–193

Schmelter T (2007b) Consideration on group-wise identical designs for linear mixed models. J Stat Plan Inference 137:4003–4010

Schmelter T, Benda N, Schwabe R (2007) Some curiosities in optimal designs for random slopes. In: Jesus L-F, Juan MR-D, Bernard T (eds) Moda 8—advances in model-oriented design and analysis. Physica-Verlag, Heidelberg, pp 189–195

Schwabe R, Schmelter T (2008) On optimal designs in random intercept models. Tatra Mt Math Publ 39:145–153

Silvey SD (1980) Optimal design. Chapman and Hall, London

Sun Y, Sinha BK, Von Rosen D, Meng QY (2003) Nonnegative estimation of variance components in multivariate unbalanced mixed linear models with two variance components. J Stat Plan Inference 115:215–234

Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer series in statistics. Springer, New York

Vonesh EF, Chinchilli VM (1997) Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, New York

Wang CM, Lam CT (1997) A mixed-effects model for the analysis of circular measurements. Technometrics 39:119–126